一种基于进化算法的概化理论最佳样本量估计新方法：兼与三种传统方法比较*

黎光明[1] 秦越[1,2]

（[1]华南师范大学心理学院、心理应用研究中心，广州 510631）（[2]广州海洋地质调查局，广州 511466）

摘 要 概化理论在心理与教育测量领域应用较广。如何使测量程序在预算限制的情况下达到较优的可靠性是研究者需要考虑的重要问题，这个问题可以转换为最佳样本量估计的问题。提出了一种基于进化算法的估计概化理论下最佳样本量的新方法——约束进化算法，并采用模拟研究的方法比较了微分优化法、拉格朗日法、柯西不等式法等三种传统方法与约束进化算法的优劣。结果表明：在两侧面交叉设计、两侧面嵌套设计和三侧面交叉设计中都证明了约束进化算法更具优越性，建议研究者在今后的研究中优先使用。

关键词 概化理论，预算限制，最佳样本量估计，约束进化算法

1 引言

概化理论(Generalizability Theory)属于现代心理测验理论，广泛用于心理与教育测量领域(杨志明，张雷，2003；朱宇 等，2013；罗照盛，郭小军，2014；Truong et al., 2021)。概化理论框架的一个优点是能够精确定位测量误差的多个来源(Clayson et al., 2021)。概化理论分为概化研究(G study) 和决策研究 (D study)(漆书青 等，2002)。概化研究的目的是在明确测量目标和测量侧面的前提下，在观测全域上尽可能多的探明研究设计中的各种测量误差的来源、结构(Vispoel et al., 2020)。决策研究的目的是为了某种特殊的决策需要，以概化研究所得到的这些方差分量估计值为基础，通过调整测量过程中各方面的关系（如调整各个侧面样本水平数、调整各个侧面之间关系、改变不同变量权重等），来探索如何控制和调节测量误差(黎光明，2019a)。概化系数是指从一个测验的被试得分拓广到测验程序同等接受度的条件全域中，被试均分估计的准确性(Zhang & Lin, 2016)。概化系数被定义为全域分数方差与其和观察分数方差两者之和的比率，计算表达式为：

$$E\rho^2 = \sigma^2(\tau)/[\sigma^2(\tau) + \sigma^2(\delta)] \quad (1)$$

在公式（1）中，$E\rho^2$表示概化系数，$\sigma^2(\tau)$表示全域分数的方差分量，也就是测量目标的方差，而$\sigma^2(\delta)$表示相对误差方差，是所有与测量目标相关的测量侧面交互作用所得方差之和。概化系数的高低代表了测验可靠性的高低。由公式（1）可知，当分子——代表测量目标的方差$\sigma^2(\tau)$增大，或者代表相对误差方差的$\sigma^2(\delta)$减小时，概化系数$E\rho^2$会相应上升。也就是说，当测量误差相对变小时，概化系数$E\rho^2$会变大。

在概化理论中，随着侧面水平数量的增加，概化系数会随之提高。然而，如果在研究过程中存在限制条件，比如受人力、物力、财力等所限，那么就需要权衡是否需要对研究设计做出改变。在某些情况下，让概化系数提高所要增加的某一侧面的观察数量较大，这时需要的经费可能超出预算。当出现这种情况时，就需要权衡增加侧面观察数量的必要性(Brennan, 2001)。由此看来，预算和成本是进行测量研究时不可忽略的问题，预算的高低在某种程度上会影响测量结果的正确性。在预算限制下，找到一个高可靠性的测验程序是研究者关注的主要问题之一。

增加样本量是减少误差方差最简单也是最直观的方法。但是，样本量大小受到经费和测量内容的限制，不可能无限增大。并且由于测量设计的不同，采用单纯增加测量侧面样本数的方

收稿日期：2021–09–15

*广东省自然科学基金面上项目 (2021A1515012516)。

通信作者：秦越，E-mail：15975588450@163.com。

法增加概化系数，很可能出现即便实际费用超出预算，也得不到理想的概化系数的可能性(Marcoulides, 1993)。Cronbach 等人(1972)指出，增加测量侧面的数量，一般会使测量的概化系数提高。然而，这种提高是不稳定的，在某些条件下，增加大量的测量侧面数量也只能对概化系数带来非常微小的改变。所以，如何寻找预算与可靠性之间的平衡点，如何在总体预算限制条件下得到最优的概化系数成为研究者需要考虑的问题。另外，不同的概化设计下，不同侧面的样本量所需的成本可能是不同的，这又使如何平衡预算与概化系数的问题变得更加复杂。因此，采用怎样的方法在预算限制下估计最优概化系数，并取得最优概化系数对应的最佳样本量，成为概化理论实际测量应用中的重要问题。

多年以来，研究者对概化理论预算限制下估计最佳样本量的问题进行了一些研究(Marcoulides, 1993; Meyer et al., 2014; 黎光明, 陈子豪 等, 2020; 刘颖 等, 2020)。但是，前人的研究主要集中在传统方法领域，影响较大的有微分优化法(differential optimization method)、拉格朗日法(Lagrange method)、柯西-施瓦茨不等式法(Cauchy Schwartz inequality method，简称柯西不等式法)等。这些传统方法基本是从数学规划理论出发，通过适用性研究，得到较优解。但迄今为止，各种传统方法都存在计算复杂、适用性差、使用条件苛刻等问题，在情况较为复杂时，并不大适合在实际研究中指导最佳样本量的选择。

随着高性能计算机算法的日益发展，计算机进化算法在最优化领域已经足以与传统方法比肩，很多性能指标上可能已经超越传统方法。而在该问题上，还没有相关的理论研究和实践。这些新的问题依然有待于研究、发现和推广。基于此，本研究提出一种基于进化算法的解决该问题的新方法——约束进化算法(Constrained Optimization Evolutionary Algorithms)，并采用模拟研究的方法比较了微分优化法、拉格朗日法、柯西不等式法等三种传统方法与约束进化算法的优劣。

2 概化理论预算限制下最佳样本量估计

2. 1 传统方法

在总体成本有限、预算限制情境下寻找最佳测量设计的问题最早由 Cleary 和 Linn(1969)提出，他们发现在单侧面概化设计中，在成本允许范围内使用最多的项目数量即可获得最优概化系数，但是这种方法不具有推广性，在测量设计包含的侧面大于两个时该算法即失效了。

Woodward 和 Joe(1973)在两侧面和三侧面交叉设计条件下采用微分优化法推导出在有限预算下计算最佳样本量取值的数学方法。该方法是将数学上称之为约束条件的预算限制表达式代入概化系数计算方程后微分求极值，筛选出合理结果后返回代入约束公式，求得最终的最佳样本量。这种方法原理简单，但 Woodward 和 Joe(1973)只讨论了在简单的预算限制公式下，两侧面和三侧面条件下交叉设计的情况，并且由于原理的限制，该方法不能简单推广到其他嵌套设计中，对于更多侧面的情况，公式推导难度和计算复杂度更是呈指数级增加，所以有较大的局限性。

Macrolides 和 Goldstein(1990)利用两侧面和三侧面完全交叉随机设计的案例，首次提出了采用拉格朗日法求解预算限制下最佳样本量取值，并且首次采用放缩的方式简化了三侧面完全交叉随机模型的计算难度。

拉格朗日法的主要计算特点是引入了拉格朗日乘子λ。郑芳英和高雪芬(2018)构建了拉格朗日法一般式：

$$L(x,y,\lambda)=f(x,y)-\lambda\varphi(x,y) \qquad (2)$$

拉格朗日法求解过程为，依据公式（2），使函数$f(x,y)$在函数$\varphi(x,y)$限制下取得最优解，且对$L(x,y,\lambda)$中x、y、λ分别求偏导，并令对应一阶偏导数等于零，得到：

$$\begin{cases} L'_x=f'_x(x,y)-\lambda\varphi'_x(x,y)=0 \\ L'_y=f'_y(x,y)-\lambda\varphi'_y(x,y)=0 \\ L'_\lambda=\varphi(x,y)=0 \end{cases} \qquad (3)$$

解以上方程组，就可得到x、y、λ的值，此时解(x, y)即为待优化函数$f(x,y)$在限制条件$\varphi(x,y)$下取到的极值，若极值点只有一个，则为该待优化问题的最优解。在概化理论预算限制下最佳样本量估计问题中，一般将含有样本自变量的相对误差方差函数σ_δ^2作为待优化函数，将预算限制公式的变形公式作为限制条件，最终求得最佳样本量取值。

但是，Macrolides 和 Goldstein(1990)只给出了拉格朗日法的计算方法，但没有论证这种方法的适用性，也缺乏阐述其运行原理。随后，Goldstein 和 Marcoulides(1991)从原理上论证了拉格朗日法在三侧面随机模型上求得预算限制下最佳样本量取值的正确性。同时，提出了一种二分搜索查找法，用一个三侧面随机交叉设计实例演示了该方法的运用，但这种方法步骤琐碎，难以应用。并且，Goldstein 和 Marcoulides(1991)提到了一个较为粗糙的拉格朗日法在随机四侧面模型下解决预算限制问题的初步公式，以表示拉格朗日法更具广泛的适用性，但这组公式较为复杂，也没有给出实证数据，且依然缺乏拉格朗日法在其他概化设计中适用性的研究。Marcoulides(1993)首次对多种预算限制条件进行讨论，给出了几种适用于不同实际测量情境的不同预算限制表达式。Meyer 等人(2014)通过一个学生评教的实例，检验了拉格朗日法在两侧面以及三侧面多种嵌套设计下对预算限制下样本量求解问题的适用性。但是，拉格朗日法还是存在适应性不佳、对计算要求高等问题。

Stuart(1954)在抽样调查领域提出了利用柯西-施瓦茨不等式估计带约束条件下最佳样本量取值的方法。他提到，在抽样调查问题中总样本方差等于各侧面方差与样本量比值的加和，即$v = \sum(v_h/n_h)$，其中v为总样本方差，v_h为计算所得侧面h的样本方差分量，n_h为侧面h的样本量。但是，他提出的计算方法并不能简单应用于概化理论中。Sanders(1992)首次将柯西不等式法应用于求解概化理论预算限制下最佳样本量，提出了一个随机两侧面设计在几种预算表达式下的最佳样本量求解方法，并给出了一个两侧面的实例。

柯西不等式的基本定义为：设有两组任意实数a_1、$a_2 \dots a_n$和b_1、$b_2 \dots b_n$，则不等式：

$$\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 \geq \left(\sum_{i=1}^{n} a_i b_i\right)^2 \quad (4)$$

当且仅当$\frac{a_1}{b_1} = \frac{a_2}{b_2} = \cdots = \frac{a_n}{b_n} = k$时（$k$为常数），公式（4）成立。Sanders(1992)为了将这种方法引入概化理论预算限制下最佳样本量求解问题中，将预算表达式定义为各侧面数量与每个侧面固定费用乘积之和，即$c = \sum c_h n_h$，其中c为总预算，c_h为侧面h单位样本的固定费用，n_h为侧面h的样本量。仿照柯西不等式的基础公式，可以得到：

$$v = \left(\sum \frac{v_h}{n_h}\right)\left(\sum c_h n_h\right) \geq \left(\sum \sqrt{v_h c_h}\right)^2 \quad (5)$$

当且仅当$\sqrt{\frac{v_h}{n_h^2 c_h}} = k$（$k$为一个常数）时，公式（5）成立。由于$k$为解出的常数，而$v_h$、$c_h$、$c$、$v$等参数已知，故可以由此解出最佳样本量$n_h$的值。由于需要满足柯西不等式的定义，因此每种概化设计都需要单独构造计算公式，且柯西不等式严格要求预算表达式呈加法形式。

在计算上，柯西不等式法难度适中，但需要构造特殊不等式，所以增加了相当程度的构造难度，并严格依赖于加法的预算表达式，故较难推广。

2.2 传统方法预算限制条件

在概化理论预算限制下最佳样本量估计的研究中，先后出现了多种预算限制表达式。Woodward 和 Joe(1973)提出了简单的样本量乘积的限制表达式$n_i n_j = L$，其中n_i、n_j为侧面样本量，L为总预算限制。Marcoulides(1993)提出了根据单位成本定义的连乘式预算表达式$c n_1 n_2 \dots n_n \leq B$，其中$c$为完成测验每一项目的单位成本，$n_1$、$n_2 \dots n_n$为各侧面的水平数，$B$为总预算。为了满足部分研究中不同侧面所需费用不一致，或同侧面水平费用不一致的情况，研究者又陆续提出了多种多项式形式的预算表达式，如$c_1 n_1 + c_2 n_2 + c_{12} n_1 n_2 \leq B$、$c_1 n_1 + c_2 n_2 \leq B$、$c_1 n_1 + c_{12} n_1 n_2 \leq B$等，其中$c_1$、$c_2$、$c_{12}$为符合某种预算定义的单位费用，$n_1$、$n_2$为对

应侧面的样本量，B为总预算。

在概化理论预算限制下最佳样本量估计的传统方法中，预算表达式的选择是一个不可忽视的问题。预算限制条件可以是简单的，也可以是复杂的，取决于研究设计和设置(Jutkowitz et al., 2019)：一方面，不同的概化设计可能对预算表达式有特殊的要求，比如可能出现同侧面不同水平花费不一致的现象，或者需要固定某一侧面花费等特殊情况，这就需要预算表达式做出相应的改变；另一方面，不同的预算限制下最佳样本量求解方法对预算表达式也有特殊的要求，如微分优化法依赖于复杂度低的预算表达式，一旦预算表达式过于复杂，可能出现无法收敛或者计算复杂度过高的问题，而柯西不等式法则依赖于连加形式的预算表达式，无法对连乘形式的预算表达式进行运算。

2.3 新方法

概化理论预算约束条件下最佳样本量估计的问题，在数学上可以归纳为带约束最优化问题(Constrained Optimization Problems)。近年来，依靠高性能计算发展起来的进化算法逐渐脱颖而出，成为了解决带约束最优化问题的最新选择。传统的带约束最优化方法通常只基于梯度信息，如简约梯度法、投影梯度法、内点惩罚函数法等(李笠 等, 2021)。这些方法只能优化部分问题，对复杂的非凸的、不连续不可导的、多目标多约束条件等问题毫无办法。并且，这些方法灵活性差，往往求解问题稍稍变化，就会导致方法效用降低甚至失灵。

约束进化算法通过模拟生物界基因交叉、基因变异、基因重组等方式驱动计算机算子进行进化，从而引导种群趋向最优化。约束进化算法经过多年的发展，从最初常常存在早熟收敛、搜索停滞、求解精度低等问题，发展到如今全局搜索与局部搜索并重，精度高、收敛效果好，已经相当成熟。另外，约束进化算法还具有天然并行性、不需要额外信息、输出结果多样等传统的带约束最优化方法不具备的优点。约束进化算法种类有经典遗传算法、差分进化算法、改进型差分进化算法、蚁群算法、多目标粒子群算法等(刁兴春 等, 2017)。不同的算法性能、适合问题各不相同，各有优缺点。就目前而言，在概化理论领域，尚未有约束进化算法适用性的相关研究。

本研究根据概化理论预算限制下最佳样本量估计问题的特点，选择了在单目标全局优化领域较为成熟、性能较好的“改进型差分进化算法”。该方法是一种新型、高效的启发式并行搜索技术(Neri & Tirronen, 2010)，具有优化效率高、参数设置简单、鲁棒性较好等优点(丁青锋, 尹晓宇, 2017)。改进型差分进化算法基于三种典型遗传操作驱动计算机算子不断进化，其算法的核心是将初始种群中选出的随机一对个体加权之后的差向量作为第三个个体的变异数，然后经过随机的参数混合实现交叉变异，检测适应度，进行筛选迭代，优胜劣汰，最终得到最优解。改进型差分进化算法的基本步骤如下：（1）选择初始解空间种群；（2）检测是否满足最终解条件，如果达到，程序停止；（3）对当前种群适应度等指标进行分析记录；（4）选择差向量，得到差分变异个体；（5）差分变异个体与种群进行交叉变异，得到变异种群；（6）初始种群与变异种群根据适应度等条件进行选择；（7）回到步骤 2 继续执行。

3 p×i×r 设计模拟研究

本研究采用在实践中应用较为普遍的两侧面交叉设计 p×i×r，使用蒙特卡洛数据模拟研究对四种估计方法进行比较。

3.1 p×i×r 设计概化系数及相对误差

在 p×i×r 概化设计中，概化系数及相对误差表达式为：

$$E\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{pr}^2}{n_r} + \frac{\sigma_{res}^2}{n_i n_r}} \quad (6)$$

$$\sigma_\delta^2 = \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{pr}^2}{n_r} + \frac{\sigma_{res}^2}{n_i n_r} \quad (7)$$

3.2 p×i×r 设计预算表达式

在本研究中，柯西不等式法因算法构造的需要，必须采用加法多项式形式的预算表达式，为了方便，在两侧面设计中将统一采用：

$$c_1 n_1 + c_{12} n_1 n_2 \leq B \tag{8}$$

公式(8)为预算表达式。为了不失一般性且具有实践意义，本研究选用了黎光明和欧旭伦(2020)关于评教问卷问题的费用参数作为预算表达式的费用参数，如表 1 所示。

表 1 问卷费用

项目	单价	单位成本
评教问卷工本费	1 元/份	
被试费	8 元/份	c_{12} = 0.4 元
评教问卷邮递费	1 元/份	

在表 1 中，每题单位成本c_{12} = 0.4 元，问卷设计阶段额外费用c_1 = 5 元/道。此情境下，n_1代表题目数量，n_2代表被试数量。因为本研究的目的是探讨最优化方法的正确性，所以省去预算单位，仅取c_{12} = 0.4、c_1 = 5，进行抽象研究。

3.3 p×i×r 设计最佳样本量估计方法

3.3.1 微分优化法

在 p×i×r 概化设计中，运用微分优化法求解概化设计最佳样本量，与上述方法不同，需要将预算表达式公式（8）代入到概化系数具体表达式公式（6）中，得到表达式：

$$E\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{pr}^2}{\frac{B-c_i n_i}{c n_i}} + \frac{\sigma_{res}^2}{\frac{B-c_i n_i}{c}}} \tag{9}$$

然后，将表达式（9）求偏导，得到表达式：

$$\frac{dE\rho^2}{dn_i} = -\sigma_p^2 \left[\sigma_p^2 + \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{pr}^2}{\frac{B-c_i n_i}{c n_i}} + \frac{\sigma_{res}^2}{\frac{B-c_i n_i}{c}}\right]^{-2} \cdot \left[\sigma_p^2 + \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{pr}^2}{\frac{B-c_i n_i}{c n_i}} + \frac{\sigma_{res}^2}{\frac{B-c_i n_i}{c}}\right]' \tag{10}$$

经过计算，得到最终的最佳样本量n_i与n_r的结果表达式为：

$$n_i = \frac{-2B\cdot\sigma_{pi}^2\cdot c_i n_i \pm \sqrt{\left(2B\cdot\sigma_{pi}^2\cdot c_i n_i\right)^2 - 4\left[B\cdot c\cdot\sigma_{pr}^2 + c_i\cdot c\cdot\sigma_{res}^2 - \sigma_{pi}^2\cdot {c_i}^2\right]\cdot\sigma_{pi}^2\cdot B^2}}{2\left[B\cdot c\cdot\sigma_{pr}^2 + c_i\cdot c\cdot\sigma_{res}^2 - \sigma_{pi}^2\cdot {c_i}^2\right]}, \quad n_r = \frac{B - c_i n_i}{c n_i} \tag{11}$$

3.3.2 拉格朗日法

在 p×i×r 概化设计中，运用拉格朗日法，可以将相对误差表达式公式（7），以及相应的预算表达式公式（8）代入到本文提到的拉格朗日一般式公式（2）中，得到表达式：

$$minL(n_i, n_r, \lambda) = \sigma_\delta^2 - \lambda(c_i n_i + c n_i n_r - B) \tag{12}$$

结合公式（7）与公式（8），运用上文提到的拉格朗日法求解过程，参考黎光明和欧旭伦(2020)研究中的计算方法，得到在 p×i×r 概化设计下最佳样本量n_i与n_r的表达式：

$$n_i = \frac{B}{c_i + c n_r}, \quad n_r = \sqrt{\left(\frac{\sigma_{pr}^2}{\sigma_{pi}^2}\frac{B}{c}\right) + \left(\frac{\sigma_{res}^2}{\sigma_{pi}^2}\frac{c_i}{c}\right)} \tag{13}$$

3.3.3 柯西不等式法

在 p×i×r 概化设计中，运用柯西不等式法求解最佳样本量，与拉格朗日法类似，运用相对误差表达式公式（7），以及相应的预算表达式公式（8），利用柯西不等式原理，结合 Sanders(1992)提到的概化理论柯西不等式构造法，求出柯西不等式法常数k，并结合常数k与预算表达式公式（8），参考黎光明和欧旭伦(2020)研究中的计算方法，得到公式：

$$k = \frac{\sqrt{\sigma_{pi}^2}\sqrt{c_i} + \sqrt{\sigma_{res}^2 + \frac{\sigma_{pr}^2 B}{c_i}}\sqrt{c}}{B} \tag{14}$$

运算后得到最佳样本量n_i与n_r的结果表达式：

$$n_i = \sqrt{\frac{\sigma_{pi}^2}{c_i k^2}}, \quad n_r = \frac{B - c_i n_i}{c n_i} \tag{15}$$

3.3.4 约束进化算法

与传统方法依赖计算公式不同，约束进化算法无需推导计算表达式，在进化算法完成后，只需要将约束条件的预算表达式公式（8）与待优化问题公式（7）代入到进化算法程序中，然后根据实际情况调整算法参数，即可得到最终最优解，也就是最佳样本量n_i与n_r。运用本研究所采用的改进型差分进化算法时，需调整的关键参数有：

（1）初始化目标维数 M = 1，决策变量维数 Dim = 2。

（2）编写优化目标即 pop.ObjV = evpr / x1 + evpi / x2 + evpir / (x1 * x2)。

（3）编写约束条件即 pop.CV = np.hstack([(c_1* x1 * x2+c*x2)-B])。

（4）调试进化种群适用度、变异率、重组率等重要参数。

```
myAlgorithm.MAXGEN = 10000 # 最大进化代数；
myAlgorithm.mutOper.F = 0.00001 # 差分进化中的参数 F；
myAlgorithm.recOper.XOVR = 0.7 # 重组概率；
……
```

约束进化算法是一种概率算法，种群规模、编码方式、种群适应度、变异率等参数影响该算法的效果，需要进行一定的手工调整，以达到最优效果。

3.4 计算方法研究工具

为了配合模拟研究大批量研究数据，本研究将上述计算方法完成了代码化。采用 R 语言，分别为微分优化法、拉格朗日法、柯西不等式法编写分析程序。采用 Python 语言，利用由华南农业大学、暨南大学等高校共同开发维护的 Python 进化算法工具包 Geatpy2，为改进型差分进化算法编写分析程序，并保留了程序接口，方便与模拟数据对接。

3.5 模拟数据生成

一般来说，参考 Brennan(2001)研究的流程，蒙特卡洛模拟技术在概化理论生成数据的主要过程是：

首先，将要做数据模拟的概化设计转化为相应的数学模型，以 p×i×r 设计为例，数学模型为：

$$X_{pir} = \mu + (\mu_p - \mu) + (\mu_i - \mu) + (\mu_r - \mu) + (\mu_{pi} - \mu_p - \mu_i + \mu) + (\mu_{pr} - \mu_p - \mu_r + \mu) + (\mu_{ir} - \mu_i - \mu_r + \mu) + (X_{pir} - \mu_{pi} - \mu_{pr} - \mu_{ir} + \mu_p + \mu_i + \mu_r - \mu) \tag{16}$$

其次，将数学模型转换为方差分量模式：

$$X_{pir} = \mu + \nu_p + \nu_i + \nu_r + \nu_{pi} + \nu_{pr} + \nu_{ir} + \nu_{pir,e} \tag{17}$$

再次，将方差分量模式转换为：

$$X_{pir} = \mu + \sigma_p z_p + \sigma_i z_i + \sigma_r z_r + \sigma_{pi} z_{pi} + \sigma_{pr} z_{pr} + \sigma_{ir} z_{ir} + \sigma_{pir} z_{pir} \tag{18}$$

其中，σ_p、σ_i、σ_r、σ_{pi}、σ_{pr}、σ_{ir}、σ_{pir}等为已知的方差参数，μ为分数的期望，在模拟研究中一般设定为 0，z_p、z_i、z_r、z_{pi}、z_{pr}、z_{ir}、z_{pir}一般设定为介于（0，1）之间的随机数，利用计算机语言中的正态随机数生成函数生成（R 语言中为 rnorm 函数）。

最后，将所有分量连加连乘，得到模拟数据X_{pir}，在实际使用中，需要根据各侧面的样本数量确定数据矩阵的格式和大小，形成最终的模拟数据矩阵。

本研究在蒙特卡洛模拟数据生成的过程中使用的编程语言为 R 语言。因为 R 语言在模拟数据过程中运算效率不高，只能单核单线程序运行，在处理大规模矩阵运算、大批量数据模拟中是十分耗时的，所以本研究采用了 Docker 容器虚拟化技术，将蒙特卡洛模拟程序运行在数台 Linux 高性能服务器中，实现了模拟程序多核并行计算。相较于普通 PC 机，程序的运行效率提高了 20 倍以上。这种技术将来也可以在其他模拟研究中使用。

3.6 模拟数据参数估计

在利用 R 语言蒙特卡洛模拟程序语句生成模拟数据矩阵后，本研究采用 R 语言中较为成熟的基于混合线性模型和限制性极大似然法概化理论软件包 gtheory 进行参数估计。

对于 p×i×r 的数学模型，根据上文提到的模拟数据生成方式，需要确定模拟数据参数σ_p、σ_i、σ_r、σ_{pi}、σ_{pr}、σ_{ir}、σ_{pir}等。为了模拟研究的真实性且不失一般性，采用了黎光明(2019b)在研究中提到的两侧面交叉设计方差分量作为本次模拟研究的参数(出于统一性的考虑，更换了部分符号)，如表 2 所示。

表 2 p×i×r 设计方差分量

	p	i	r	pi	pr	ir	pir, e
方差分量	0.01009	0.01289	0.00047	0.02070	0.00189	0.01130	0.03256

根据实践中的研究经验，本设计将测量目标 p 的样本量n_p固定为 30。为了探讨在测量侧面不同水平数情况下各预算限制最佳样本量估计方法的有效性，在模拟数据生成中，本研究采用了多种测量侧面水平数，此处选取n_i =（20, 40），n_r = （5, 10），形成了 2×2 四种不同的模拟设计。为了配合四种设计，从实践出发，本研究也将预算限制分为四档 B =（100, 150, 200, 250）。因为数据量较大，每种设计模拟 500 次，共 2000 次。四批数据生成后，利用上文提到的 gtheory 包分析方差分量，为了后续分析，将所得 500×4 个结果分别输出到 csv 文件中。

3.7 模拟研究结果

为了便于分析，分别对得到的 4×4（四种模拟设计×四种方法）组最佳概化系数取平均数，通过比较概化系数平均数来比较四种方法，得到的结果如表 3 所示。

3.8 分析与讨论

根据表 3，在 p×i×r 概化设计中，四种算法在四种样本量条件下得到的预算限制下最佳概化系数都很优异，都可以满足人们日常研究的需求。值得注意的是，三种传统方法在该条件下，所得到的分析结果几乎相同，这可能与计算机程序的精度和估算方法有关，这也从侧面证明了本研究关于微分优化法基于新的预算表达式的推导是正确的。

约束进化算法在 p×i×r 概化设计下，得到的概化系数与三种传统方法相差无几，并且在多次比较中，都比传统方法更好地将花费限制在了预算范围内。

在运算时间上，传统方法因为所有的分析过程都是在计算机程序之外完成的，所以看起来比约束进化算法快一些。但是，考虑到实际研究中，概化设计方案的选择仅在测验研究时进行一次，这点时间几乎可以忽略不计。

表 3 四种计算方法比较结果

条件组合	样本量组合									
	20 × 5					20 × 10				
	概化系数	最佳样本量n_i	最佳样本量n_r	设计花费	分析用时/s	概化系数	最佳样本量n_i	最佳样本量n_r	设计花费	分析用时/s
微分优化法	0.819773870	13.1667055	6.4872856	100	0.002	0.85970836	18.9712839	7.26671699	150	0.0060
	0.814549990	13	6	96.2		0.85844574	19	7	148.2	
拉格朗日法	0.819773870	13.1667055	6.4872856	100	0.006	0.85970836	18.9712839	7.26671699	150	0.0079
	0.814549990	13	6	96.2		0.85844574	19	7	148.2	
柯西不等式法	0.819773870	13.1667055	6.4872856	100	0.006	0.85970836	18.9712839	7.26671699	150	0.0078
	0.814549990	13	6	96.2		0.85844574	19	7	148.2	
约束进化算法	0.818306752	14.0492881	5.2942692	99.9	9.795	0.85935208	18.5389829	7.67551280	149.9	10.072
	0.814821502	14	5	98		0.86328631	19	8	155.8	

表 3（续） 四种计算方法比较结果

条件组合	样本量组合									
	40 × 5					40 × 10				
	概化系数	最佳样本量n_i	最佳样本量n_r	设计花费	分析用时/s	概化系数	最佳样本量n_i	最佳样本量n_r	设计花费	分析用时/s
微分优化法	0.889393370	24.28057970	8.09258899	200	0.0034	0.906841124	29.4206630	8.74357291	250	0.00350
	0.888094949	24	8	196.8		0.906658958	29	9	249.4	
拉格朗日法	0.889393370	24.28057970	8.09258899	200	0.0027	0.906841124	29.4206630	8.74357291	250	0.00223
	0.888094949	24	8	196.8		0.906658958	29	9	249.4	
柯西不等式法	0.889393370	24.28057970	8.09258899	200	0.0025	0.906841124	29.4206630	8.74357291	250	0.00454
	0.888094949	24	8	196.8		0.906658958	29	9	249.4	
约束进化算法	0.888315630	24.01884929	8.03969713	200	9.9540	0.906071456	31.1473389	7.49944160	250	9.75300
	0.888094949	24	8	196.8		0.903522912	31	7	241.8	

4（r:p）×i 设计模拟研究

本研究进行了两侧面嵌套设计（r:p）×i 的模拟研究，其数学推导与 p×i×r 设计相似，结果表明，四种算法的表现与 p×i×r 设计中的表现大体相同，限于篇幅，不再赘述。

5 p×i×r×o 设计模拟研究

5.1 p×i×r×o 设计概化系数及相对误差

三侧面交叉设计 p×i×r×o 是经典的三侧面概化设计，也是其他同类概化设计的基础。例如，在实践中，p 可以代表被试，i 可以代表测验题目，而 r 可以代表评分员，o 可以代表不同的测验场合。

p×i×r×o 三侧面交叉设计概化系数及相对误差表达式：

$$E\rho^2 = \frac{\sigma_p^2}{\sigma_p^2+\frac{\sigma_{pi}^2}{n_i}+\frac{\sigma_{pr}^2}{n_r}+\frac{\sigma_{po}^2}{n_o}+\frac{\sigma_{pir}^2}{n_in_r}+\frac{\sigma_{pio}^2}{n_in_o}+\frac{\sigma_{por}^2}{n_rn_o}+\frac{\sigma_{res}^2}{n_in_rn_o}} \tag{19}$$

$$\sigma_\delta^2 = \frac{\sigma_{pi}^2}{n_i}+\frac{\sigma_{pr}^2}{n_r}+\frac{\sigma_{po}^2}{n_o}+\frac{\sigma_{pir}^2}{n_in_r}+\frac{\sigma_{pio}^2}{n_in_o}+\frac{\sigma_{por}^2}{n_rn_o}+\frac{\sigma_{res}^2}{n_in_rn_o} \tag{20}$$

5.2 p×i×r×o 设计预算表达式

在概化理论预算限制下最佳样本量估计的研究中，三侧面设计推导与计算难度远大于两侧面，导致各种计算方法对预算表达式的要求变得苛刻起来。在三侧面设计下，微分优化法和拉格朗日法都要求使用连乘式预算表达式：

$$cn_1n_2n_3 \le B \tag{21}$$

而柯西不等式法在三侧面情况下则依然要求多项式式预算表达式：

$$c_1n_1 + c_2n_2 + c_3n_3 \le B \tag{22}$$

只有约束进化算法可以兼容所有类型的预算表达式，所以在三侧面设计下，本研究以约束进化算法为桥梁，分两次比较四种方法的优劣。

为了便于模拟研究并且不失代表性和普遍性，本研究假设连乘预算表达式中$c=1$，多项式表达式中$c_1=20$、$c_2=30$、$c_3=40$。

5.3 p×i×r×o 设计最佳样本量估计方法

5.3.1 微分优化法

在 p×i×r×o 概化设计中，运用微分优化法求解概化设计最佳样本量，要求较为苛刻，Woodward 和 Joe(1973)提出了一种复杂的六次多项式表达式解法，是将形如公式（21）的连乘预算表达式代入三侧面概化系数表达式公式（19）中，并求偏微分，得到n_r表达式：

$$0 = An_r{}^6 + Bn_r{}^5 + Cn_r{}^4 + Dn_r{}^3 + En_r{}^2 + Fn_r + G \tag{23}$$

$$A = \sigma_{pio}^4\sigma_{pi}^2\sigma_{po}^2L$$

$$B = \sigma_{pio}^4\sigma_{por}^2\sigma_{pi}^2L + \sigma_{pio}^4\sigma_{pir}^2\sigma_{po}^2L - \sigma_{po}^4\sigma_{pi}^4L^2$$

$$C = \sigma_{pio}^4\sigma_{por}^2\sigma_{pir}^2L + 2\sigma_{pr}^2\sigma_{pio}^2\sigma_{po}^2\sigma_{pi}^2L^2$$

$$D = 2\sigma_{po}^2\sigma_{pir}^2\sigma_{pi}^2\sigma_{por}^2L^2 - 2\sigma_{pr}^2\sigma_{pio}^2\sigma_{por}^2\sigma_{pi}^2L^2 - 2\sigma_{pr}^2\sigma_{pio}^2\sigma_{pir}^2\sigma_{po}^2L^2$$

$$E = \sigma_{pr}^4\sigma_{po}^2\sigma_{pi}^2L^3 - 2\sigma_{pr}^2\sigma_{pio}^2\sigma_{por}^2\sigma_{pir}^2L^2$$

$$F = \sigma_{por}^2\sigma_{pi}^2\sigma_{pr}^4L^3 - \sigma_{pir}^4\sigma_{por}^4L^2 + \sigma_{po}^2\sigma_{pir}^2\sigma_{pr}^2L^3$$

$$G = \sigma_{por}^2\sigma_{pir}^2\sigma_{pr}^4L^3$$

解出n_r的值后，再根据表达式：

$$n_i = \sqrt{\frac{n_r\sigma_{pi}^2L+\sigma_{pir}^2L}{\sigma_{po}^2n_r{}^2+\sigma_{pio}^2n_r}}\ ,\quad n_o = \frac{L}{n_in_r} \tag{24}$$

解出n_i、n_o的值。

5.3.2 拉格朗日法

针对三侧面最佳样本量估计计算困难的问题，Marcoulides 和 Goldstein(1990)提出了一种调整相对误差方差表达式的方法，相当于数学上的放缩法，以获得相对误差方差的上界。本研究对三侧面最佳样本量估计的拉格朗日法也采用了这种放缩法。

首先，将三侧面相对误差表达式公式（20），放缩为：

$$\sigma_\delta^2 = \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{pr}^2}{n_r} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{pir}^2}{n_i} + \frac{\sigma_{pio}^2}{n_o} + \frac{\sigma_{por}^2}{n_r} + \frac{\sigma_{res}^2}{n_i n_r n_o} \tag{25}$$

其次，将预算表达式公式（21）代入到本文提到的拉格朗日一般式公式（2）中，得到：

$$minL(n_i, n_r, n_o, \lambda) = \sigma_\delta^2 - \lambda(c n_i n_r n_o - B) \tag{26}$$

最后，参考 Marcoulides 和 Goldstein(1990)的结论，得到在 p×i×r×o 概化设计下最佳样本量n_i、n_r、n_o的表达式为：

$$n_i = \sqrt[3]{\frac{(\sigma_{pi}^2+\sigma_{pir}^2)^2}{(\sigma_{po}^2+\sigma_{pio}^2)(\sigma_{pr}^2+\sigma_{por}^2)}}\left(\frac{\bar{c}}{c}\right),\ n_r = \sqrt[3]{\frac{(\sigma_{pr}^2+\sigma_{por}^2)^2}{(\sigma_{po}^2+\sigma_{pio}^2)(\sigma_{pi}^2+\sigma_{pir}^2)}}\left(\frac{\bar{c}}{c}\right),\ n_o = \sqrt[3]{\frac{(\sigma_{po}^2+\sigma_{pio}^2)^2}{(\sigma_{pi}^2+\sigma_{pir}^2)(\sigma_{pr}^2+\sigma_{por}^2)}}\left(\frac{\bar{c}}{c}\right) \tag{27}$$

在公式（27）中，$\frac{\bar{c}}{c}$ 在本研究中相当于B。

5.3.3 柯西不等式法

在前人的研究中，并没有将柯西不等式运用到三侧面设计中的研究，其原因可能与柯西不等式法构造较为复杂、较难推广有关。本研究利用放缩法，将柯西不等式法推广到三侧面交叉设计，如下：

首先，将 p×i×r×o 设计的相对误差表达式公式（20），放缩为：

$$\sigma_\delta^2 = \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{pr}^2}{n_r} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{pir}^2}{n_i} + \frac{\sigma_{pio}^2}{n_o} + \frac{\sigma_{por}^2}{n_r} + \frac{\sigma_{res}^2}{n_i} \tag{28}$$

其次，参照公式（22），构建预算表达式：

$$c_i n_i + c_o n_o + c_r n_r \le B \tag{29}$$

最后，根据柯西不等式的构造法，可得n_i、n_r、n_o的表达式：

$$n_i = \frac{B}{c_i + c_o\sqrt{\frac{(\sigma_{po}^2+\sigma_{pio}^2)\cdot c_i}{(\sigma_{pi}^2+\sigma_{pir}^2+\sigma_{res}^2)\cdot c_o}} + c_r\sqrt{\frac{(\sigma_{pr}^2+\sigma_{por}^2)\cdot c_i}{(\sigma_{pi}^2+\sigma_{pir}^2+\sigma_{res}^2)\cdot c_r}}},\ n_r = \sqrt{\frac{(\sigma_{pr}^2+\sigma_{por}^2)\cdot c_i}{(\sigma_{pi}^2+\sigma_{pir}^2+\sigma_{res}^2)\cdot c_r}}\cdot n_i,\ n_o = \sqrt{\frac{(\sigma_{po}^2+\sigma_{pio}^2)\cdot c_i}{(\sigma_{pi}^2+\sigma_{pir}^2+\sigma_{res}^2)\cdot c_o}}\cdot n_i \tag{30}$$

5.3.4 约束进化算法

在三侧面设计研究中，约束进化算法依然无需推导计算表达式。因为约束进化算法可以兼容公式（21）和公式（22）两种预算表达式，所以，约束进化算法在这里起到了比较桥梁的作用，即三种传统方法分别可与约束进化算法进行比较。运用本研究所采用的改进型差分进化算法时，需调整的关键参数有：

（1）初始化目标维数 M = 1，决策变量维数 Dim = 2。

（2）编写优化目标：pop.ObjV=evc_p/(evc_p+evc_pi/x1)+evc_pj/x2+evc_pk/x3、+evc_pij/(x1*x2)+evc_pik/(x1*x3)+evc_pjk/(x2*x3)+evc_res/(x1*x2*x3))。

（3）编写两种约束条件即 pop.CV = np.hstack([(c1*x1+c2*x2+c3*x3)-B]) 以及 pop.CV = np.hstack([x1*x2*x3 -B])。

（4）调试进化种群适用度、变异率、重组率等重要参数。

myAlgorithm.MAXGEN = 40000 # 最大进化代数;

myAlgorithm.mutOper.F = 0.00001 # 差分进化中的参数 F;

myAlgorithm.recOper.XOVR = 0.8 # 重组概率;

……

5.4 计算方法研究工具

p×i×r×o 研究与 p×i×r 研究所使用的计算方法研究工具类似，不再赘述。

5.5 模拟数据生成

根据 p×i×r×o 的数学模型：

$$X_{piro} = \mu + \nu_p + \nu_i + \nu_r + \nu_o + \nu_{po} + \nu_{io} + \nu_{ro} + \nu_{pi} + \nu_{pr} + \nu_{ir} + \nu_{pir} + \nu_{pio} + \nu_{pro} + \nu_{iro} + \nu_{piro,e} \quad (31)$$

根据上文提到的模拟数据生成方式，需要确定模拟数据参数σ_p、σ_i、σ_r 、σ_o、σ_{po}、σ_{pi}、σ_{pr}、σ_{ir}、σ_{pir}…等。为了模拟研究的真实性且不失一般性，采用了 Meyer 等人(2014)在研究中提到的三侧面交叉设计方差分量作为本次模拟研究的参数(出于统一性的考虑，更换了部分符号)，如表 4 所示。

表 4 p×i×r×o 设计方差分量

	p	i	r	o	pi	po	io	io	pr
方差分量	0.227	0.013	0.192	0.005	0.026	0.085	0.006	0.006	0.053

表 4（续） p×i×r×o 设计方差分量

	io	ir	ro	pi	pio	pir	iro	pro	piro, e
方差分量	0.006	0.007	0.009	0.021	0.298	0.053	0.005	0.017	0.321

根据实践中的研究经验，本设计将测量目标 p 的样本量n_p固定为 30。为了探讨在测量侧面不同水平数情况下各预算限制最佳样本量估计方法的有效性，在模拟数据生成中，本研究采用了大小两种测量侧面水平数，n_i = (15,25)，n_r = (5,10)，n_o= (5,10)，形成了 2×2×2 八种不同的模拟设计，每种设计的预算取值分别为 B = (800,900,1000,1100,1200,1300,1400,1500)。因为数据量特别大，所以每种设计仅模拟 200 次，共 1600 次。八批数据生成后，利用上文提到的 gtheory 包分析方差分量，为了后续分析，将所得 200×8 个结果分别输出到 csv 文件中。

5.6 模拟数据参数估计

得到八批方差分量数据后，利用上文 5.3 提到的四种计算方法以及自编 R 语言微分优化法、拉格朗日法、柯西不等式法分析程序，以及 Python 语言约束进化算法分析程序，进行批量分析，得到 200×8 组概化系数、最佳样本量、设计花费、分析用时等结果，并将结果输出至 csv 文件，方便后续分析。

5.7 模拟研究结果

为了便于分析，本研究分别对得到的 5×8（五种方法×八种模拟设计，其中约束进化算法出现两次）批次最佳概化系数取平均数，通过比较概化系数平均数比较四种方法，出于简洁的原因，这里只呈现 5×4 种结果（如表 5 和表 6 所示）。

表 5 约束进化算法与微分优化法、拉格朗日法比较

条件组合	样本量组合											
	5×15×5						5×15×10					
	概化系数	最佳样本量n_i	最佳样本量n_o	最佳样本量n_r	设计花费	分析用时/s	概化系数	最佳样本量n_i	最佳样本量n_o	最佳样本量n_r	设计花费	分析用时/s
微分优化法	0.87817495	2.3127595	82.202016	4.2080120	800.0	0.012	0.88471850	2.5147869	80.646650	4.4376697	900	0.022
	0.86539109	2	82	4	656		0.88810806	3	81	4	972	
拉格朗日法	0.92365199	5.6781953	27.620465	6.6991252	1050.7	0.004	0.92652447	5.9837807	28.622702	6.9407586	1188.8	0.006
	0.92704140	6	28	7	1176		0.92713459	7	34	8	1218	
约束进化算法	0.88528957	5.7857628	49.674564	2.7834917	799.9	5.207	0.91616891	11.913346	10.781614	6.9123128	1000	5.218
	0.89205920	6	50	3	900		0.91748451	12	11	7	924	

表 5（续） 约束进化算法与微分优化法、拉格朗日法比较

条件组合	样本量组合											
	5×25×5						5×25×10					
	概化系数	最佳样本量n_i	最佳样本量n_o	最佳样本量n_r	设计花费	分析用时/s	概化系数	最佳样本量n_i	最佳样本量n_o	最佳样本量n_r	设计花费	分析用时/s
微分优化法	0.89293562	2.682893	80.032004	4.657285	1000	0.032	0.89277702	2.7866442	79.717925	4.9517096	1100	0.041
	0.90227221	3	80	5	1200		0.89713009	3	80	5	1200	
拉格朗日法	0.93024553	6.187979	29.937401	7.132465	1321.3	0.008	0.92911955	6.3653751	30.705416	7.4224466	1450.7	0.009
	0.92881883	6	30	7	1260		0.92562861	6	31	7	1302	
约束进化算法	0.92295047	4.401688	23.866680	9.518668	994.8	5.203	0.92338934	6.1970158	23.784237	7.4631270	1098	5.350
	0.92057362	4	24	10	960		0.92043541	6	24	7	1008	

表 6 约束进化算法与柯西不等式法比较

条件组合	样本量组合											
	5×15×5						5×15×10					
	概化系数	最佳样本量n_i	最佳样本量n_o	最佳样本量n_r	设计花费	分析用时/s	概化系数	最佳样本量n_i	最佳样本量n_o	最佳样本量n_r	设计花费	分析用时/s
柯西不等式法	0.902348495	14.329990	11.437680	4.256743	800.0	0.006	0.913370488	16.1628300	12.8614720	4.7724801	900.0	0.008
	0.897368063	14	11	4	770		0.915588593	93	64	20	910	
约束进化算法	0.903530483	8.553932	13.441316	4.896444	770.2	5.230	0.916469323	7.7272937	13.5470196	7.4085236	857.2	5.420
	0.904627504	9	13	5	770		0.916700553	8	14	37	860	

表 6（续） 约束进化算法与柯西不等式法比较

条件组合	样本量组合											
	5×25×5						5×25×10					
	概化系数	最佳样本量n_i	最佳样本量n_o	最佳样本量n_r	设计花费	分析用时/s	概化系数	最佳样本量n_i	最佳样本量n_o	最佳样本量n_r	设计花费	分析用时/s
柯西不等式法	0.923125551	17.938602	14.342051	5.2741606	1000.00	0.008000135	0.926730356	19.7373660	15.731487	5.8327017	1100.0	0.00999999
	0.920326101	18	14	5	980		0.928381049	20	16	6	1120	
约束进化算法	0.92431176	9.8778782	12.819470	8.2053678	910.35	5.330000000	0.926673922	8.2434067	17.705147	8.2420120	1025.7	5.43000000
	0.92435965	10	13	8	910		0.925507605	8	18	8	1020	

5.8 分析与讨论

由表 5 可以看出，在 p×i×r×o 三侧面概化设计中，拉格朗日法虽然得到的概化系数更高，但代价是普遍地大幅度超出预算限制，导致这种方法的指导意义严重下降。而微分优化法虽然严格执行了预算限制，但由于优化公式较为笨拙，无法达到较为优秀的优化结果，而约束进化算法无论在对预算的遵守上，还是在优化结果上，都表现得非常优秀。

由表 6 可以看出，在 p×i×r×o 三侧面概化设计中，柯西不等式法也能取到较为优秀的概化系数，但需要注意的是，柯西不等式法严格依赖特定的预算表达式，若预算表达式稍有变动，则柯西不等式无法使用。然而，约束进化算法即便在符合柯西不等式特定的预算表达式下，也取得了比柯西不等式法更为优异的概化系数，表现出较强的稳健性。

约束进化算法在两种预算表达式下都表现良好，这也表明了约束进化算法具有传统方法不可比拟的扩展性和鲁棒性。柯西不等式法的效果较好，从侧面也说明了本研究首次提出的三侧面柯西不等式方法的推导是正确的。

在运算时间上，传统方法因为所有的分析过程都是在计算机程序之外完成的，所以看起来比约束进化算法快一些。但是，考虑到实际研究中，概化设计方案的选择仅在测验研究时进行一次，这点时间几乎可以忽略不计。

6 综合讨论

6.1 预算舍入的影响

从以上研究结果可以看出，四种概化理论预算限制下最佳样本量估计方法均为连续优化，所得的最佳样本量结果都为非整数。而在测量实践中，一般要求样本量为整数，可以采用四舍五入的方法取得近似最优解。然而，一方面这种舍入取得的整数样本量未必是最佳的整数样本量，另一方面，这种舍入常常会造成成本超出预算限制，这点各研究中已经被反复证实。在测验规模较大时，这时超出的预算费用可能非常可观，而这是人们应该避免的。避免因舍入导致测量费用与预算不符的一个相对简便的操作方法是估算舍入后样本可能的所有排列组合，并从中选择最佳的观察数量。另外，Saunders 等人(1989)提出了整数规划方法，也可用于获得最佳样本量。整数规划方法的优点是不需要对结果进行舍入，但这种方法可行性较差，对算法设计要求高，计算量大，其实用性不如本研究提到的四种算法。

6.2 方差分量为负数的影响

在某些概化理论研究设计中，所估计的方差分量可能为负值，这可能源于数据本身的问题。一般而言，在进行研究时，方差分量为负的情况可以当作方差分量为 0 处理。Brennan(2001)指出，将负数的方差分量代入计算或将其归零代入计算，所得结果可能无异，但需注意负数方差分量的出现代表着数据的正确性可能出现了问题。除此之外，由于传统方法需要进行大量分式运算，将方差分量直接设为 0 可能会出现无法求解的情况。然而，本研究所使用的约束进化算法则不会出现这种问题。

6.3 固定侧面的影响

本研究讨论的三种概化设计均为随机侧面概化设计。然而，在某些情况下，仅仅需要一定的侧面水平数量，或者某些侧面水平是不可变的，则这种设计就成为固定侧面概化设计。在固定侧面概化设计中，当固定侧面数量增加时，测验的误差来源就会减少，相应的概化系数就会提高。在概化理论预算限制下最佳样本量估计的研究中，Meyer 等人(2014)讨论了关于含有固定侧面的三侧面概化设计的求解方法。显然，含有固定侧面的问题相对较为简单，可以由随机侧面的最佳样本量求解问题简化解决。

6.4 不平衡设计的影响

在概化理论测量实践应用中，有可能出现不平衡设计的情况。在本研究中，（r:p）×i 设计为典型的不平衡设计，这是因为侧面 p 的各水平数不一致。一般地，在概化理论方差分量估计中，这种情况应该当作缺失数据处理。然而，在概化理论中，对缺失数据的处理方法是将缺失值以最大的水平数为标准进行空值补齐，不平衡数据中不同水平样本量差别很大时，需要进行空值补齐的缺失值就会非常多，使数据总量发生巨大变化，从而对方差分量估计产生巨大影响，可能偏离真实情况。然而，本研究所使用的约束进化算法有效地解决了不平衡设计的影响。

6.5 计算机程序对传统方法的影响

本研究所涉内容众多、数据复杂，所以采用了自编程序的方法，将三种传统方法程序化进行批量计算求解。虽然有计算迅速、使用方便等优点，但不可忽视的是，计算机本身的精度限制和计算原理的限制可能对计算结果产生微小影响。例如，应用求解方程的计算机方法，本研究微分优化法在将某些方差分量代入计算时，可能会出现找不到解的情况，虽然最终采用了较为复杂的方法解决了这个问题，但仍然值得警惕和注意。本研究对三种传统方法的程序化使用的是 R 语言，但可能会因其他计算机语言的计算精度或原理与 R 语言的不同而产生细微差别。

6.6 放缩法对传统方法的影响

为了满足三侧面概化设计求解预算限制下最佳样本量的要求，Marcoulides 和 Goldstein(1990)提出了一种调整相对误差方差表达式的方法，相当于数学上的“放缩法”，以获得相对误差方差的上界，他将这种放缩思想带入了拉格朗日法求解过程。本研究也使用了这种“放缩法”，将柯西不等式的适用范围推广到了三侧面概化设计中。但是，在模拟研究中都可以看出，使用“放缩法”的拉格朗日法虽然能够得到比较好的概化系数，但是超出预算限制比较明显。在预算总额不高的情况下，超出部分可能不明显，但在高预算的大型测验中，这种程度的超出预算则是很难被接受的。本研究首次提出在三侧面概化设计中对柯西不等式法进行放缩应用，在特定预算表达式条件下得到了较好的结果，并符合预算。但是，需要注意的是，由于放缩法原理的限制，必然会产生不精确的结果，所以在应用中，应特别注意在所得结果的一定范围内搜寻最优解。

6.7 约束进化算法性能的影响

约束进化算法作为计算机程序算法，所使用的计算机性能会影响其运行速度。本研究为了探讨一般情况下的约束进化算法的使用情况，没有使用在模拟研究中使用的 Linux 高性能工作站进行计算，而是采用了普通 PC（i5 处理器，8G 内存）在 Windows 环境下进行运算。得到的运算时间都在秒级，运算速度完全能够满足实际研究的需要。另外，本研究所使用的改进型差分进化算法，是约束进化算法中对单目标优化较为成熟的一种。但是，约束进化算法种类繁多，仅在单目标优化领域就有多染色体遗传算法、多染色体稳态遗传算法等很多种新算法，每种算法都有各自特点，不排除其他算法在概化理论预算限制下最佳样本量求解问题上有更优异的表现。

7 结论

通过将三种传统方法与约束进化算法在 p×i×r 设计、（r:p）×i 设计和 p×i×r×o 设计的模拟研究进行比较，本研究发现：

（1）在两侧面概化设计问题的优化性能上，约束进化算法稍好于传统方法，而三种传统方法表现相当。在两侧面 p×i×r 设计与（r:p）×i 设计下，当预算表达式相同时，三种传统方法所得到最佳样本量结果几乎一致，都表现出了很好的性能。但不约而同的是，在某

些条件下会超出预算。而约束进化算法不仅在精度上达到了与传统方法相同，而且在对预算限制条件的遵守上更佳。

（2）在三侧面概化设计问题的优化性能上，约束进化算法明显好于传统方法，三种传统方法表现存在差异。在 p×i×r×o 设计下，微分优化法得到的概化系数最差，但对预算的遵守性最好；拉格朗日法得到的概化系数最好，但超出预算较多，在对预算要求较为严格时无法形成有效指导建议。在与微分优化法和拉格朗日法比较中，约束进化算法不仅取到了相对最优的概化系数，还较好地遵守了预算限制，综合表现明显优于其他两种方法。在约束进化算法与柯西不等式法的比较中，柯西不等式法在特殊预算表达式表现较好，既得到了较好的概化系数，又较好地遵守了预算限制，而约束进化算法的综合表现则稍好于柯西不等式法，尤其更加贴近预算限制。

（3）在复杂度上，约束进化算法明显低于传统方法。三种传统方法复杂度均比较高，在两侧面设计下尚可正常推导和计算，在三侧面设计下就变得非常困难，在四侧面及以上设计下则几乎不可推导和计算。并且，传统方法在多侧面上的推广上严重依赖于放缩法，但放缩法可能易使算法精度降低，导致侧面越多，传统方法性能越差。而约束进化算法 简洁，复杂度远低于传统方法。约束进化算法在各种设计下计算方法几乎相同，只需要经验操作就可以得到非常好的结果，对于普通研究者非常友好。

（4）在适用性上，约束进化算法显著优于传统方法。三种传统方法的适用性均比较狭窄，拉格朗日法适用性相较于其他两种传统方法稍好。但是，三种传统方法均依赖于特定的概化设计和预算表达式，当概化设计或预算表达式发生变化时，则可能无法满足需求。而约束进化算法则不依赖特定的概化设计与预算表达式，在两侧面、三侧面等概化设计和不同预算表达式下均表现良好，适用性极强。

（5）在推广性上，约束进化算法明显好于传统方法。三种传方算法由于数学原理的限制，难以推广到更加复杂的概化设计上，并在计算上可行性相对较差。而约束进化算法则具有广泛的扩展性，可以推广到四侧面及更复杂的概化设计中。使用多目标约束进化算法还可以推广到多元概化设计的最佳样本量估计问题上，应用空间广阔。

（6）在得到最优解的可能性上，约束进化算法也优于传统方法。因为进化算法是一种概率算法，所以本研究中一次运行得到的结果几乎不一定是最优解，若多次运行，则能找到更佳的样本量估计结果，故在某些条件下，约束进化算法可以发现传统方法无法发现的最优解。

（7）综合以上结论，本研究认为，在概化理论预算限制下最佳样本量估计问题上，约束进化算法作为一种新方法，优于其他传统方法，建议研究者在今后的研究中优先使用。

参考文献

Brennan, R. L. (2001). *Generalizability theory.* New York, NY: Springer.

Clayson, P. E., Carbine, K. A., Baldwin, S. A., Olsen, J. A., & Larson, M. J. (2021). Using generalizability theory and the erp reliability analysis (era) toolbox for assessing test-retest reliability of erp scores part 1: Algorithms, framework, and implementation. *International Journal of Psychophysiology*, *166*(1), 174–187.

Cleary, T. A., & Linn, R. L. (1969). Error of measurement and the power of a statistical test. *British Journal of Mathematical and Statistical Psychology, 22*(1), 49–55.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: theory of generalizability for scores and profiles. *American Educational Research Journal, 11*(1), 1–23.

Diao, X., Liu, Y., Cao, J., & Shang, Y. (2017). Reviews of multiobjective ant colony optimization. *Computer Science, 44*(10), 7–13.

[刁兴春, 刘艺, 曹建军, 尚玉玲. (2017). 多目标蚁群优化研究综述. *计算机科学*, *44*(10), 7–13.]

Ding, Q., & Yin, X. (2017). Research survey of differential evolution algorithms. *CAAI Transactions on Intelligent Systems*, *12*(4), 431–442.

[丁青锋, 尹晓宇. (2017). 差分进化算法综述. *智能系统学报*, *12*(4), 431–442.]

Goldstein, Z., & Marcoulides, G. A. (1991). Maximizing the coefficient of generalizability in decision studies. *Educational & Psychological Measurement*, *51*(1), 79–88.

Jutkowitz, E., Alarid-Escudero, F., Kuntz, K. M., & Jalal, H. (2019). The curve of optimal sample size (coss): A graphical representation of the optimal sample size from a value of information analysis. *Pharmaco Economics*. *37*(7), 871–877.

Li, G. (2019a). *Psychological measurement*. Beijing, China: Tsinghua Universiy Publishing House.

[黎光明. (2019a). *心理测量*. 北京: 清华大学出版社.]

Li, G. (2019b). Estimation of optimal sample size under budget constraints for generalization theory based on LaGrange multiplier method. *Statistics & Decision*, *527*(11), 13–16.

[黎光明. (2019b). 基于拉格朗日乘法的概化理论预算限制下最佳样本量估计. *统计与决策*, *527* (11), 13–16.]

Li, G., Chen, Z., & Zhang, M. (2020). Estimating the best sample size of teaching level evaluation for college teachers under budget constraints in generalizability theory. *Psychological Development and Education*, *36*(3), 378–384.

[黎光明, 陈子豪, 张敏强. (2020). 高校教师教学水平评价概化理论预算限制下最佳样本量估计. *心理发展与教育*, *36*(3), 378–384.]

Li, G., & Ou, X. (2020). Comparing of LaGrange multiplication and Cauchy-Schwarz inequality for optimal sample size estimation. *Statistics & Decision*, *552*(12), 29–33.

[黎光明, 欧旭伦. (2020). 拉格朗日乘法与柯西不等式最佳样本量估计比较. *统计与决策*, *552*(12), 29–33.]

Li, L., Li, G., Chang, L., & Gu, T. (2021). Survey of constrained evolutionary algorithms and their applications. *Computer Science*, *48*(4), 1–13.

[李笠, 李广鹏, 常亮, 古天龙. (2021). 约束进化算法及其应用研究综述. *计算机科学*, *48*(4), 1–13.]

Liu, Y., Zhang, M., & Zhen, F. (2020). Estimating the best sample size for students' evolution of teaching——Based on the application of LaGrange multiplier method. *Journal of Psychological Science*, *43*(4), 857–863.

[刘颖, 张敏强, 甄锋泉. (2020). 学生评教研究的最佳样本量估算——基于拉格朗日乘数法的应用. *心理科学*, *43*(4), 857–863.]

Luo, Z., & Guo, X. (2014). The optimal size of material in psychological experiment: The applications of multivariate generalizability theory. *Acta Psychologica Sinica*, *46*(6), 876−884.

[罗照盛, 郭小军. (2014). 认知行为实验研究中最佳素材容量的选择与确定：多元概化理论应用. *心理学报*, *46*(6), 876–884.]

Marcoulides. G. A. (1993). Maximizing power in generalizability studies under budget constraints. *Journal of Educational Statistics*, *18*(2), 197–206.

Marcoulides, G. A., & Goldstein, Z. (1990). The optimization of generalizab-ility studies with resource constraints. *Educational & Psychological Measurement*, *50*(4), 761–768.

Meyer, P. J., Liu, X., & Mashburn, A. J. (2014). A practical solution to optimizing the reliability of teaching observation measures under budget constraints. *Educational and Psychological Measurement*, *74*(2), 280–291.

Neri, F., & Tirronen, V. (2010). Recent advances in differential evolution: a survey and experimental analysis. *Artificial Intelligence Review, 33*(1-2), 61–106.

Qi, S., Dai, H., & Ding, S. (2002). *Modern educational and psychological measurement*. Beijing, China : Higher Education Press.

[漆书青, 戴海崎, 丁树良. (2002). *现代教育与心理测量学原理.* 北京:高等教育出版社.]

Sanders, P. F. (1992). Alternative solutions for optimization problems in generalizability theory. *Psychometrika, 57*(3), 351–356.

Saunders, P. F., Theunissen, T. J., & Baas, S. M. (1989). Minimizing the number of observations: A generalization of the Spearman Brown formula. *Psychometrika, 54*(4), 587–589.

Stuart, A. (1954). A simple presentation of optimum sampling results. *Journal of the Royal Statistical Society. Series B: Methodological, B16*(2), 239–241.

Truong, Q. C., Choo, C., Numbers, K., Merkin, A. G., Brodaty, H., Kochan, N. A., Sachdev, P. S., Feigin,V. L., & Medvedev, O. N. (2021). Applying generalizability theory to examine assessments of subjective cognitive complaints: whose reports should we rely on – participant versus informant?. *International Psychogeriatrics, 4*, 1-11.

Vispoel, W. P., Xu, G., & Kilinc, M., (2020). Expanding G-Theory models to incorporate congeneric relationships: Illustrations using the big five inventory. *Journal of Personality Assessment, 103*(1), 1–14.

Woodward, J., & Joe, G. (1973). Maximizing the coefficient of generalizability in multi-facet decision studies. *Psychometrika, 38*(2), 173–181.

Yang, Z., & Chang, L. (2003). *Generalizability theory and its applications*. Beijing, China: Educational Science Publishing House.

[杨志明, 张雷. (2003). *测评的概化理论及其应用.* 北京: 教育科学出版社.]

Zhang, J., & Lin, C. K. (2016). Generalizability theory with one-facet nonadditive models. *Applied Psychological Measurement, 40*(6), 367–386.

Zheng, F., & Gao, X. (2018). A sufficient condition for conditional extreme value in using lagrange multiplier method. *Studies in College Mathematics, 21*(2), 41–43.

[郑芳英, 高雪芬. (2018). 拉格朗日乘子法求条件极值的充分条件. *高等数学研究, 21*(2), 41–43.]

Zhu, Y., Fung, S., & Xin, T. (2013). Improving dependability of new HSK writing test Scores: A generalizability theroy based approach. *Journal of Psychological Science, 36*(2), 479–488.

[朱宇, 冯瑞龙, 辛涛. (2013). 新 HSK 书写成绩可靠性影响因素的概化理论分析. *心理科学, 36*(2), 479–488.]

A New Method for Estimating the Optimal Sample Size in Generalization Theory Based on Evolutionary Algorithm: Comparisons with Three Traditional Methods

LI Guangming[1], QIN Yue[1,2]

([1]School of Psychology, Center for Studies of Psychological Application, South China Normal University, Guangzhou 510631, China)

([2]Guangzhou Marine Geological Survey, Guangzhou 511466, China)

Abstract

Generalizability Theory (GT) is widely applied in psychological measurement and evaluation. A larger generalizability coefficient often indicates a higher reliability the test may have. Generalizability coefficients can be improved by increasing sample sizes. However, the size of a sample would be subject to budget constraints. Therefore, it is important to examine how to effectively determine the size of a sample considering the budget

constraints. The existing literature has been largely limited to traditional methods, such as the differential optimization method (Woodward & Joe, 1973), the Lagrange method (Macrolides & Goldstein, 1990), and the Cauchy Schwartz inequality method (Sanders, 1992).

These traditional methods have limited scope of application and their typical conditions are hard to satisfy. In addition, there is no unified comparison available. Fortunately, with the increased use of high performance computing, the Constrained Optimization Evolutionary Algorithms (COEAs) becomes highly feasible.

This paper expands and compares the four methods—the differential optimization method, Lagrange method, Cauchy Schwartz inequality method, and COEAs—determine the best solution to the optimal sample size problem under the budget constraints in GT. Specifically, this paper compares the applicability of the four methods using three generalizability designs, including p× i ×r, (r: p) ×i and p ×i ×r ×o designs. The results are presented as follows:

(1) In the optimization performance of two-facet generalizability design of p× i × r and (r: p) × i, the performance of COEAs is slightly better than that of the traditional methods, whereas the performance of three traditional methods is equivalent. Although COEAs and the traditional methods have showed similar accuracy, the former has better compliance concerning budget constraints.

(2) In the optimization performance of three-facet generalizability design of p ×i ×r ×o, the performance of COEAs is obviously better than that of the traditional methods. The least ideal generalizability coefficient is obtained using the differential optimization method, whereas its budget compliance is the best; the generalizability coefficient obtained by Lagrange method is the best, but higher than the budget. The Cauchy inequality method obtains a better generalizability coefficient under special budget constraints. But, the performance of COEAs is slightly better than that of Cauchy Schwartz inequality method, especially closer to the budget constraints.

(3) In terms of the algorithm complexity, COEAs obtains an obviously smaller algorithm complexity than do the traditional methods. The complexity of the three traditional methods is relatively high. However, COEAs does not rely on the derivation of mathematical formulas, and the algorithm is relatively less complex.

(4) In terms of the algorithm applicability, COEAs is significantly better than the traditional methods. The applicability of the three traditional methods is relatively narrow. However, COEAs does not rely on a specific generalizability design or a budget expression, and, therefore, the applicability of COEAs is stronger.

(5) In terms of the algorithm generalizability, COEAs is obviously better than the traditional methods. The limited mathematical principles make it difficult to extend the three traditional methods to more complex generalizability designs, and thus, the feasibility of calculation is poor. Howerve, COEAs has revealed stronger generalizability.

(6) In terms of the possibility of getting the best solution, COEAs is also better than the traditional methods. Because evolutionary algorithm is a probabilistic algorithm, multiple tests can be conducted to obtain better results for optimal sample sizes. Under some conditions, COEAs can determine better solutions, which, however, is impossible for three traditional methods.

(7) These results suggest that COEAs is superior to three traditional methods in estimating the optimal sample size problem under the budget constraints in GT. It is recommended that researchers use COEAs in future research to determine an optimal sample size in their psychological measurement and evaluation.

Key words Generalizability Theory, budget constraints, sample size estimation methods, Constrained Optimization Evolutionary Algorithms

黎光明：提出研究思路，设计研究方案；负责最终版本修订。
秦越：负责进行实验；负责数据模拟和分析数据；负责论文起草。